#So Berlin?

Where to move within Germany, if you want to enjoy the night while caring about rental prices and wages?

Clustering Germany's biggest cities according to their nightlife, average income and rental prices

Felix Klapprott

Mai, 2020

Part I: Introduction

1.1. Background

Have you ever been to Berlin? If yes - you might have heard someone saying "omg that's so Berlin". Berlin has become one of the most popular places to live for young people, not only from Germany but from all around the world. This is partly because of its vibrant nightlife and the diversity of potential nightly activities. It also used to be a very affordable place to live (and still is compared to other European capitals). While for some people, a couple of big famous clubs with their almost unlimited freedom are the main attractions, many appreciate the options of going to clubs as well as to pubs or cocktail bars. This attractiveness came with a price. As Berlin became more popular, rental prices have gone up and entry requirements for many of Berlin's universities became stricter. Therefore other cities like Leipzig benefited from Berlin losing its appeal as city which is "arm aber sexy" (poor but sexy; (in)famous quote by a former mayor). The question where to study and consequently where to move, is often based on the city rather than on the university. Hence many students and young professionals wonder where to move to account for both living on a financially sound basis and yet hedonistically enjoy themselves.

1.2. Problem

After leaving school many young people in Germany choose a new place to live, in order to study and start their own life. However choosing a city to live in is not always easy.

There are many young people in Germany who focus on nightlife-options when choosing a place to live, but typically they also know that a city should provide the opportunity to earn enough money to pay their rent. Different cities in Germany come with a certain set of stereotypes regarding how "fun" they are, whether rental costs are high and also whether it is likely to earn a high wage in a given city.

Typically they are not based on data and therefor little is known about a city's attractiveness for young people.

Using a cluster analysis, I'd like to make it easier to compare different cities, so that young people get an easier impression whether less famous cities (e.g. Nuremburg, Hannover, Bremen) in Germany could be equally attractive compared to big established cities such as Berlin or Munich.

Foursquare data about the diversity or pattern of potential nightlife activities in a city can be used to determine which cities resemble each other regarding their nightlife in order to analyse options for people seeking to relocate. As even the most hedonistic night owl needs to eat and sleep at some place, accounting for rental prices and average income is useful.

1.3. Interest

Young people who plan to move to another city and which are strongly driven by enjoying a given style of nightlife might be interested.
Especially when choosing a city to enrol in a Bachelor's or Master's degree young adults in Germany typically make their choice based on a city and not so much based on a certain university. In order to make an informed decision and to rely less on stereotypes young adults could be interested in such an analysis.
For example: Imagine someone who has just finished their Bachelors at one city in Germany and now wonders where to apply for their Masters within Germany. A data driven approach might show additional options and could bring his or her attention to cities previously which were ignored.

Additionally, as students are an attractive target group for cities and municipalities, to bring future skilled labour to their area; city marketing departments might also be interested in which cities are main competitors in that regard and what they could highlight in an effort to bring young talents to their cities.

Part II: Data

For my analyses, I will use data on Income, Rental Prices and a city's nightlife.

I decided to focus the analyses on cities with more than 500 000 Inhabitants. A useful overview is found freely available on the German Wikipedia at https://de.wikipedia.org/wiki/Liste_der_Gro%C3%9F-_und_Mittelst%C3%A4dte_in_Deutschland .
An extracted list will be the basis of all analyses as it provides a list of official names and the needed selection of all German cities with more than 500.000. The list will be used to find the names of all cities with more than 500.000 Inhabitants in order to find out more about them in later steps (i.e. using official statistics (rental prices / income), geo-data (visualization) and foursquare data (nightlife))

The foursquare API will provide me with useful information about the type of venues in a given city. I will retrieve information about relevant venues for all of the cities with more than 500.000 inhabitants. I will focus only on venues which are registered as “Nightlife spot” and plan to look for a specific pattern of different venues rather than the mere number of options. I will calculate relative frequencies of different venue types per city to find similarities between cities, i.e. for example distinguishing cities where you would find mainly cocktail and wine bars as opposed to night clubs, pubs and beer gardens.

Additionally rental prices for these cities are needed. Rental prices will be used as a variable in the later cluster analysis to find patterns among Germanys biggest cities (i.e. are their overlooked and remarkably cheap yet interesting cities in Germany?). The data set (German) including a comparison to 2009 is accessible (premium/university access) at https://de.statista.com/statistik/daten/studie/167163/umfrage/mietentwicklung-in-den-deutschen-grossstaedten/ . As it is not publicly available I will download it and convert to csv in order to work with it in Python. The data is based on 234,000 offers advertised on the online platform immowelt.de. The prices reflect the median of the rental apartments and houses offered in the first half 2019. As many students decide to share flats to benefit from old contracts with cheaper prices these data is only an approximation of true costs of living. Nevertheless it will give an impression of whether a city is rather expensive or cheap to live in.

Average Income will be used as a variable in the later cluster analysis to find patterns among Germanys biggest cities and to distinguish cities based on economic potential for later graduates, as this might also influence their decision. The most recent data I could find on accessible income per capita in German cities was collected in 2016. The data set can be found on https://de.statista.com/statistik/daten/studie/998971/umfrage/verfuegbares-einkommen-in-den-groessten-staedten-in-deutschland/ . As it is not easily available I will download it too and convert to csv in order to work with it in Python.

Part III Methodology

Choosing the cities of interest

I had to focus my analyses on a number of cities. I decided to use a cut-off of 500.000 Inhabitants, as typically larger cities are more attractive for young adults.

I used the package BeautifulSoup to parse the Wikipedia-Page 'https://de.wikipedia.org/wiki/Liste_der_Gro%C3%9F-_und_Mittelst%C3%A4dte_in_Deutschland', where all German cities with more than 20.000 Inhabitants and their respective size is listed. After some additional cleaning (i.e. removing problematic signs and changing the datatype to float), I removed all cities with less then 500.000 Inhabitants, leaving me with a dataset of 14 Cities between 518.365 (Nuremburg) and 3.644.826 Inhabitants (Berlin).

Extracting additional non-foursquare data

I downloaded the official data from statista.de, using my student access, manually extracted relevant columns and uploaded the data set on github to allow access for fellow coursera-students.

I then imported the csv-data using the wget-library into python. Note that latin1-enconding was used to deal with German Umlaute such as Ö or Ü. As the name of Frankfurt am Main (francfort) was stored differently in the data set, I had to change one of the entries from “Frankfurt” to “Frankfurt am Main” in order to allow further calculations.

Using the Normatim geolocator, I added geodata (latitude/longitude) for each of the 14 cities to allow later visualization and retrieval of relevant venues in the cities proximity.

Extracting foursquare data

Using the foursquare-API I accessed venues of category “Nightlife spot” (internal category ID 4d4b7105d754a06376d81259) within a radius of 5km surrounding the respective city center. This was done embedding the json-request in a for loop to retrieve relevant venues not only for one, but for all of the 14 cities.

Using dummy coding, I extracted the venue subcategory (i.e. Beer garden) of each individual entry and calculated relative frequency for each subcategory per city, as this was the data needed for later analyses.

Clustering

To prepare clustering, I merged data from income statistics, rental prices and relative frequencies of venue subcategories. Using K-Means clustering and arbitrarily choosing 5 clusters I assigned the 14 cities to five clusters, based on the several variables.

Visualization was done using Folium.Map in order to show the cities distribution while at the same time accounting for their assigned cluster.

Further exploratory inspection of the different cluster was done by grouping the venue frequency for each cluster to find common nightlife patterns typical for the cluster as opposed to for each city individually. As rental prices and income data was easy to inspect visually I decided against aggregation and merged city-data (prices/income) and cluster data (nightlife pattern) instead.

Part IV Results & Discussion

I found five clusters of different size with strong external validity (as judged by someone living roughly 30 years in Germany).

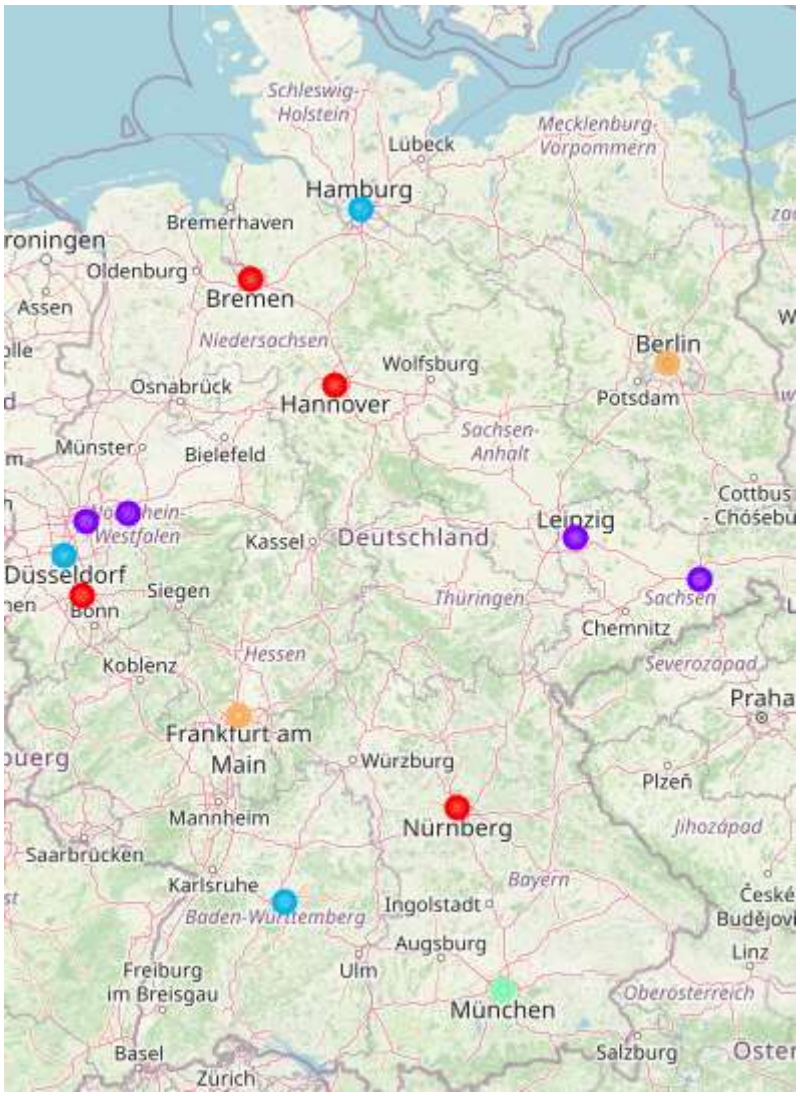

Figure 1 Visualization of five clusters (German cities > 500k ; based on Income, Rental Prices and Nightlife Structure)

The first cluster comprises the four cities Leipzig, Dresden, Dortmund and Essen. Located in the relative far east and far west of the country, they share relatively low average income (17.7k to 20.2k) and low rental prices (6.8 Euro/sqm to 7.6 Euro/sqm). Night clubs are the second most common venue for this cluster which is the highest ranking entry compared to all the other clusters. Due to

their relative low income/price structure and the absence of Wine Bars and Cocktail bars among the top three I decided to describe the cluster as “Down to earth”-Cities

The second cluster comprises the three cities Stuttgart, Hamburg, Dusseldorf. Being the state capitals of the respective federal state, they share high average income (24.4k to 25.0k) and high rental prices (10.6 Euro/sqm to 13.0/sqm). Cocktail bars are the second most common venue for this cluster, yet nightclubs are also among the top 4. Due to their relative high income/price structure and the presence of wine bars as well as cocktail bars among the top ten I decided to call the cluster “the fancy ones”.

The third cluster comprises only one city, namely the Bavarian capital of Munich. Being the state capital and world famous German tourist attraction, it has a remarkable number of beer garden, German and Bavarian restaurants as well as cafes marked as nightlife spot. Income as well as rental prices are particularly high (29.7k Income; 18.6 Euro/sqm rental price) in Munich. As it contains only one city, I decided to call the cluster “Well... Munich”.

The fourth cluster comprises two cities known for international and diverse atmosphere: Francfort and Berlin. Both have in common that rental prices are relatively high (11.6 to 14.2 Euro/sqm) given the accessible income (19.7k to 21.7 k) in the two cities. While partly world famous for its nightclubs the numbers of Bars, Cocktail Bars, Wine Bars could reflect that both cities host an affluent upper class. Being a financial power house (Francfurt) or a tourist attraction (Berlin) leads to having Hotel Bars listed at Rank 10 of most common night life venues.

Finally, the fifth and last cluster comprises the four cities Nuremberg, Hanover, Cologne and Bremen. These are cities which (maybe with the exemption of cologne) often go rather unnoticed and share an average income (21.0k to 21.8k) and average rental cost structure (8.2 to 10.9 Euro/sqm). Bars, Cocktail Bars, Pubs and Nightclubs are among the most common nightlife spots there.

Part V Conclusion

Especially for those looking to find a cheap place to study and to go out dancing, It might be worth looking to the far west (Essen/Dortmund) in addition to increasingly popular cities in the far east (Dresden/Leipzig).

Hamburg, Stuttgart and Dusseldorf have a reputation as comparably rich cities. This is supported by the analyses; however one needs to be aware that strong yet small subcultures are not reflected in the data.

Interestingly while Munich is one of the cities within Germany tourists visit most frequently and where they supposedly get into contact with German culture, it is in fact rather unique given the nightlife structure and also the accessible income and the costs of rent.

Berlin has a comparably large gap between income and rental costs. Other cities seems to have more nightclubs (relatively speeking). One might conclude that Berlins local politics need to support its trademark club culture (unlike focussing on the mere number of alternatives) to preserve it‘s reputation and to remain attractive to students.

There is a group of cities which (maybe with the exemption of cologne) often goes rather unnoticed and share an average income and average rental cost structure. These cities might benefit from targeting undergraduate students in other cities within the cluster to recruit more students willing to live in less profile cities where Bars, Cocktail Bars, Pubs and Nightclubs are among the most common nightlife spots.

Limitations

The criteria I used are by now means exhaustive. Regional differences might make it hard to adapt even when moving within a cluster (Someone from Stuttgart might struggle with a nordic‘s mentality in Hamburg; as would someone from Cologne in Hannover). Within cities such as Hamburg exists a vibrant subculture which would rightfully reject the label „fancy“.

While relative frequencies are often useful – absolute numbers could be worth looking to, especially as the cities differ in size and therefore in the overall offer they can provide young adults.

However considering all this, the external validity of this clustering is good and especially often overlooked cities might benefit to get to know more about their unique selling points compared to more popular cities to attract students and young professionals.